DESCRIPTION

A Human Nuclear Protein having a WW Domain and A Polynucleotide encoding the Protein

INS A1

Technical Field

The present invention relates to a novel protein having a WW domain and existing in human cell nuclei, a polynucleotide encoding this protein, and an antibody against this protein. The protein and antibody of the present invention are useful for diagnosis and therapy of various diseases, and the polynucleotide of the present invention is useful as a probe for genetic diagnosis or as a genetic source for gene therapy. Further, the polynucleotide can be used as a genetic source for large-scale production of the protein of this invention.

Background Art

The term "nuclear protein" is a generic name of proteins functioning in cell nucleus. In nucleus there are genomic DNA serving as a plan of organism, and nuclear proteins are involved in replication, transcriptional regulation etc. of these genomic DNA. Typical nuclear proteins whose functions have been revealed include a transcription factor, a splicing factor, an intranuclear receptor, a cell cycle regulator and a tumor suppressor. These factors are closely related not only to life phenomena such as development and differentiation but also to diseases such as cancers (New Medical Science, "Tensha No Shikumi To Shikkan" (Mechanism of Transcription and Diseases) ed. by Masahiro Muramatsu). Accordingly, these nuclear proteins are expected as

target proteins for developing low-molecular pharmaceutical preparations that regulate transcription and translation of specific genes, and it is desired to obtain as many nuclear proteins as possible.

The WW domain belongs to a new family of protein-protein interaction motifs resembling SH2, SH3, PH and PTB domains. It is known that this domain consists of about 40 amino acid residues containing 2 conserved tryptophan residues, and like the SH3 domain, binds to a proline-rich amino acid sequence (H. I. Chen and M. Sudol., Proc. Natl. Sci. 92, 7819-7823, 1995). As a result of X-ray crystallographic analysis of a WW domain/ligand conjugate, it was revealed that the three-dimensional structure of the WW domain is different from that of SH3 (M. J. Macias et al., Nature, 382, 646-649, 1996). Like other protein motifs, the WW domain is contained in the cytoskeleton system (P. Bork and M. Sudol TIBS, 19, 531-533, 1994), in proteins participating in the signal transduction system (H. I. Chen and M. Sudol., Proc. Natl. Sci., 92, 7819-7823, 1995), in a ubiquitin-protein ligase in the protein degradation system (O. Staub et al., EMBO J., 15, 2371-2380, 1996) and in a transcription activator (P. Bork and M. Sudol, TIBS, 19, 531-533, 1994), and is believed to play an important role in the intracellular signal transduction system.

The object of the present invention is to provide a novel protein present in human cell nucleus, a polynucleotide encoding this protein, and an antibody against this nuclear protein.

Disclosure of Invention

To achieve the object described above, the present application provides the following inventions (1) to (7):

(1) An isolated and purified human nuclear protein comprising the amino acid sequence of SEQ ID NO: 1.

(2) A polynucleotide encoding the protein of the invention (1), which comprises the nucleotide sequence of SEQ ID NO: 2.

(3) The polynucleotide of the invention (2), consisting of the nucleotide sequence of SEQ ID NO: 2.

(4) A human genomic DNA fragment with which a polynucleotide of SEQ ID NO:3 or a partial contiguous sequence thereof hybridizes under stringent conditions.

(5) An expression vector expressing the polynucleotide of the invention (2) or (3) in *in vitro* translation or in host cells.

(6) A transformed cell producing the human nuclear protein of the invention (1), which is transformant with the expression vector of the invention (5).

(7) An antibody against the human nuclear protein of the invention (1).

Best Mode for Carrying Out the Invention

The protein of the invention (1) can be obtained by a method of isolation thereof from human organs, cell lines etc., by a method of preparing the peptide through chemical synthesis on the basis of the amino acid sequence set forth in SEQ ID NO: 1 or by a method of production thereof by recombinant DNA

technique using the polynucleotide encoding the amino acid sequence of SEQ ID NO: 1, among which the method with recombinant DNA technique is preferably used. For example, a vector harboring the polynucleotide of the invention (2) or (3) is subjected to *in vitro* transcription to prepare RNA which is then used as a template in *in vitro* translation, whereby the protein can be expressed *in vitro*. Further, by integrating the polynucleotide in a conventional method into a suitable expression vector, the protein encoded by the polynucleotide can be expressed in a large amount in procaryotes such as *E. coli*, *Bacillus subtilis* etc. or eucaryotes such as yeasts, insect cells and mammalian cells.

To produce the protein of the invention (1) by expressing the DNA through *in vitro* translation, the polynucleotide of the invention (2) or (3) is integrated in a vector harboring an RNA polymerase promoter (the invention (5)) and added the vector to an *in vitro* translation system such as a rabbit reticulocyte lysate or a wheat germ extract containing an RNA polymerase compatible with said promoter, whereby the protein of the invention (1) can be produced *in vitro*. The RNA polymerase promoter includes e.g. T7, T3 and SP6. The vector harboring such RNA polymerase promoter includes e.g. pKA1, pCDM8, pT3/T7 18, pT7/3 19, and pBluescript II.

To produce the protein of the invention (1) by expressing the DNA in microorganisms such as *E. coli*, the polynucleotide of the invention (2) or (3) is integrated in an expression vector harboring an origin capable of replication in microorganisms, a promoter, a ribosome-binding site, a DNA cloning site, a terminator etc. to prepare the expression vector (the invention (5)) which is then used for transformation of host cells, and by culturing the resulting transformant (the invention (6)), the protein encoded by said polynucleotide can be produced in a large amount in the microorganism. If an initiation codon and a termination codon have been added respectively to sites upstream and

downstream from an arbitrary translated region in said expression vector, a protein fragment containing the arbitrary region can be obtained by expressing the DNA. Alternatively, it can also be expressed as a fusion protein with another protein. By cleaving this fusion protein with a suitable protease, the part of only the protein encoded by said polynucleotide can be obtained. The expression vector for *E. coli* includes e.g. pUC series vectors, pBluescript II, pET expression system vectors and pGEX expression system vectors.

To produce the protein of the invention (1) by expressing the DNA in eucaryotes, the translated region of the polynucleotide of the invention (2) or (3) is integrated in an eucaryotic expression vector harboring a promoter, a splicing region, a poly(A)-additional site etc. to prepare the expression vector (the invention (5)) which is then used for transforming eucaryotic cells (the invention (6)), whereby the protein of the invention (1) can be produced in the eucaryotic cells. The expression vector includes e.g. pKAI, pCDM8, pSVK3, pMSG, pSVL, pBK-CMV, pBK-RSV, EBV vector, pRS and pYES2. If vectors such as pIND/V5-His, pFLAG-CMV-2, pEGFP-N1 and pEGFP-C1 are used, the protein of the present invention can also be expressed as a fusion protein having various tags such as His tag, FLAG tag and GFP added thereto. As the eucaryotic cells, mammalian cultured cells such as simian renal cells COS7 and Chinese hamster ovary cells CHO, budding yeasts, fission yeasts, silkworm cells and Xenopus oocytes are generally used, but insofar as the protein of the invention (1) can be expressed, any eucaryotic cells can be used. For introducing the expression vector into eucaryotic cells, conventional methods such as the electroporation method, calcium phosphate method, liposome method and DEAE-dextran method can be used.

For isolating and purifying the protein of the invention (1) from a culture after expression of the desired protein in the procaryotic or eucaryotic cells, separation techniques known in the art can be used in combination.

Such techniques include e.g. treatment with a denaturant such as urea or a surfactant, sonication, enzymatic digestion, salting-out or solvent precipitation, dialysis, centrifugation, ultrafiltration, gel filtration, SDS-PAGE, isoelectric focusing, ion-exchange chromatography, hydrophobic chromatography, affinity chromatography and reverse phase chromatography.

The protein of the invention (1) encompasses peptide fragments (each consisting of 5 or more amino acid residues) containing any partial amino acid sequence from the SEQ ID NO: 1. Such a peptide fragment can be used as an antigen for preparing the antibody of the present invention. Further, the protein of the invention (1) encompasses fusion proteins with another arbitrary protein. For example, fusion proteins with glutathione-S-transferase (GST) or green fluorescent protein (GFP), described in the Examples, can be mentioned.

The polynucleotide (cDNA) of the invention (2) or (3) can be cloned from a cDNA library derived from e.g. human cells. The cDNA is synthesized using poly(A)$^+$RNA as a template extracted from human cells. The human cells may be either cultured cells or cells excised by an operation etc. from the human body. The cDNA can be synthesized by any methods such as the Okayama-Berg method (Okayama, H. and Berg, P., Mol. Cell Biol., 2, 161-170, 1982) and the Gubler-Hoffman method (Gubler, U. and Hoffman, J. Gene, 25, 263-269, 1983), but for efficiently obtaining full-length clones, the Capping method (Kato, S. et al., Gene, 150, 243-250, 1994) described in the Examples is preferably used.

The polynucleotide of the invention (2) comprises the nucleotide sequence of SEQ ID NO: 2, and for example, the polynucleotide consisting of the nucleotide sequence of SEQ ID NO: 3 has a 2669-bp nucleotide sequence containing a 2115-bp open reading frame (ORF). This ORF encodes a protein consisting of 704 amino acid residues. The polynucleotide of the invention (3)

comprises the 2115-bp nucleotide sequence (SEQ ID NO:2) constituting this ORF. By expressing the cDNA of the invention (2) or (3) in *E. coli* or animal cultured cells, an about 80-kDa protein was obtained. This protein binds to a C-terminal domain of RNA polymerase II, so it is considered to participate in transcriptional regulation.

Since the protein of the invention (1) is expressed in any tissues, the same clone as the polynucleotide of the invention (2) or (3) can be easily obtained from a human cDNA library prepared from human cells by screening the library with an oligonucleotide probe synthesized on the basis of the nucleotide sequence of the polynucleotide set forth in SEQ ID NO: 2 or 3. Alternatively, the objective cDNA can also be synthesized by polymerase chain reaction (PCR) by use of such oligonucleotides as primers.

Generally, polymorphism of human genes occurs frequently due to individual variations. Accordingly, those polynucleotides where in SEQ ID NO: 2 or 3, one or more nucleotides have been added, deleted and/or substituted with other nucleotides fall under the scope of the invention (3) or (4).

Accordingly, those proteins where in SEQ ID NO: 1, one or more amino acids have been added, deleted and/or substituted with other amino acids as a result of such alterations to nucleotides also fall under the scope of the invention (1) insofar as they have the activity of a protein having the amino acid sequence of SEQ ID NO: 1.

The polynucleotide of the invention (2) or (3) encompasses DNA fragments (10 bp or more) containing any partial nucleotide sequence from the sequence of SEQ ID NO: 2 or 3. Further, DNA fragments consisting of a sense or antisense strand thereof fall under the scope of this invention. These DNA fragments can be used as probes for genetic diagnosis.

The invention (4) is concerned with a human genomic DNA fragment with which the polynucleotide of SEQ ID NO: 3 or a partial contiguous sequence thereof hybridizes under stringent conditions. As used herein, the stringent conditions are that enables specific and detectable binding between the polynucleotide of SEQ ID NO: 3 or a partial contiguous sequence thereof (30 bp or more) and chromosome-derived genomic DNA. The stringent conditions are defined in terms of salt concentration, organic solvent (e.g., formamide), temperature and other known conditions. That is, stringency is increased by a decrease in salt concentration, by an increase in organic solvent concentration, or by an increase in hybridization temperature. For example, the stringent salt concentration is usually about 750 mM or less NaCl and about 75 mM or less trisodium citrate, more preferably about 500 mM or less NaCl and about 50 mM or less trisodium citrate and most preferably about 250 mM or less NaCl and about 25 mM or less trisodium citrate. The stringent organic solvent concentration is about 35 % or more formamide, most preferably about 50 % or more formamide. The stringent temperature condition is about 30 °C or more, more preferably about 37 °C or more and most preferably about 42 °C or more. The other conditions include hybridization time, the concentration of a detergent (e.g. SDS), the presence or absence of carrier DNA, etc., and by combining these conditions, varying stringency can be established. Further, the conditions for washing after hybridization also affects stringency. The washing conditions are also defined in terms of salt concentration and temperature, and the stringency of washing is increased by a decrease in salt concentration or by an increase in temperature. For example, the stringent salt condition for washing is about 30 mM or less NaCl and about 3 mM or less trisodium citrate, most preferably about 15 mM or less NaCl and about 1.5 mM or less trisodium citrate. The stringent temperature condition for washing is about 25 °C or more, more preferably about 42 °C or more and most preferably about 68 °C or more. The genomic DNA fragment of the invention (4) can be

isolated for example by subjecting a genome library prepared from human chromosomal DNA to screening by the above stringent hybridization with said polynucleotide as a probe and subsequent washing.

The genomic DNA fragment of the invention (4) comprises expression-regulating regions (promoter/enhancer and suppressor sequences, etc.) for the region coding for the protein of the invention (1). These expression-regulating regions are useful as a material for screening a material regulating *in vivo* expression of the protein of the invention (1).

The antibody of the invention (7) can be obtained from serum in an animal immunized with the protein of the invention (1) as an antigen. The antigen used may be a peptide chemically synthesized on the basis of the amino acid sequence of SEQ ID NO: 1 or the protein expressed in the eucaryotic or procaryotic cells. Alternatively, the antibody can be prepared by introducing the above-described expression vector for eucaryotic cells through an injection or a gene gun into animal muscles or skin and then collecting serum (e.g., an invention in JP-7-313187A). As the animal, a mouse, rat, rabbit, goat, chicken or the like is used. If a hybridoma is produced by fusing myeloma cells with B cells collected from the spleen in the immunized animal, a monoclonal antibody against the protein of the invention (1) can be produced by the hybridoma.

Examples

The present invention will be described in more detail by reference to the Examples, which however are not intended to limit the scope of the present invention. Basic procedures for DNA recombination and enzymatic reaction were in accordance with those described in a literature (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, 1989). Unless otherwise

specified, the restriction enzymes and various modifying enzymes used were products of Takara Shuzo Co., Ltd. The buffer composition in each enzymatic reaction, as well as reaction conditions, was followed instructions attached to the kits. Synthesis of cDNA was conducted according to a literature (Kato, S. et al., Gene, 150, 243-250, 1994).

(i) cDNA cloning

As a result of large-scale determination of the nucleotide sequences of cDNA clones selected from a human full-length cDNA library (described in WO97/03190), clone HP03494 was obtained. This clone had a structure made of a 291-bp 5'-untranslated region, a 2115-bp ORF and a 263-bp 3'-untranslated region (SEQ ID NO: 3). The ORF encodes a protein consisting of 704 amino acid residues.

Using the amino acid sequence (SEQ ID NO: 1) of this protein, a protein database was searched, but none of known proteins had homology to this protein. Further examination of GenBank by using the nucleotide sequence of its cDNA indicated that some ESTs (e.g. Accession No. A1758365) have 90 % or more homology thereto, but they are partial sequences, so whether or not they code for the same protein as the protein of this invention cannot be judged.

Examination of motif sequences indicated that as shown in Table 1, the region of from the 43- to 78-positions has homology to WW domains. Tryptophan residues at the 49- and 72-positions and a proline residue at the 75-position are amino acid residues conserved in every known WW domain.

Table 1

Protein	Position	Amino Acid Sequence	Accession No.
Conserved Sequence		——W——G—YY-N——W—P——	
HP03494	43	ELVHAGWEKCWSRRENRPYYFNRFTNQSLWEMPVLGQHD	
Npw38	46	EGLPPSWYKVFDPSCGLPYYWNADTDLVSWLSPHDPNSV	BAA76400
Yap_Human	171	VPLPAGWEMAKTSS. GQRYFLNHIDQTTTWQDPRKAMLS	P46937
Yap_Chick-1	169	VPLPPGWEMAKTPS. GQRYFLNHIDQTTTWQDPRKAMLS	P46936
Yap_Mouse-1	156	VPLPAGWEMAKTSS. GQRYFLNHNDQTTTWQDPRKAMLS	P46938
Ned4_Mouse-1	40	SPLPPGWEERQDVL. GRTYYVNHESRRTQWKRPSPDDDL	P46935
Ned4_Human-1	218	SPLPPGWEERQDIL. GRTYYVNHESRRTQWKRPTPQDNL	P46934
Ned4_Mouse-2	196	SGLPPGWEEKQDDR. GRSYYVDHNSKTTTWSKPTMQDDP	P46935
Ned4_Human-2	375	SGLPPGWEEKQDER. GRSYYVDHNSRTTTWTKPTVQATV	P46934
Dmd_Human	3055	TSVQGPWERAISPN. KVPYYINHETQTTCWDHPKMTELY	P11532
Dmd_Mouse	3048	TSVQGPWERAISPN. KVPYYINHETQTTCWDHPKMTELY	P11531
FE65_Rat	42	SDLPAGWMRVQDTS. GTYYWHI. PTGTTQWEPPGRASPS	P46933
Msb1/Human	249	IVLPPNWKTARDPE. GKIYYYHVITRQTQWDPPTWESPG	
IQGA_Human	679	GDNNSKWVKHWVKG. GYYYYHNLETQEGGWDEPPNFVQN	P46940
FBP11-1_Mouse	1	......WTEHKSPD. GRTYYYNTETKQSTWEKPDDLKTP	U40747
FBP11-2_Mouse	36	LLSKCPWKTYKSDS. GKPYYYNSQTKESRWAKP.....	U40747

(ii) Northern blotting

Multi tissue Northern Blot (Clontech) having human tissue poly(A)$^+$RNA blotted thereon was used as an mRNA source. As the probe, an *Eco*RI-*Not*I fragment of full-length HP03494 cDNA, labeled with a radioisotope by a random primer labeling kit (Pharmacia), was used. The conditions for Northern blotting hybridization followed the protocol attached to the kit. An about 3-kb hybridization band was obtained from the heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testicle, ovary, small intestine, colon and peripheral blood, suggesting that this protein is a housekeeping one.

(iii) Protein synthesis by *in vitro* translation

A plasmid vector harboring the polynucleotide (cDNA) of this invention was used to perform *in vitro* transcription/translation by a T_NT rabbit reticulocyte lysate kit (a product of Promega). The expression product was

labeled with a radioisotope by adding [^{35}S] methionine. Any reaction was conducted according to the protocol attached to the kit. 2 μg of the plasmid was reacted at 30 °C for 90 minutes in a 25 μl reaction solution containing 12.5 μl T_NT rabbit reticulocyte lysate, 0.5 μl buffer (attached to the kit), 2 μl amino acid mixture (not containing methionine), 2 μl (0.35 MBq/μl) of [^{35}S] methionine (Amersham), 0.5 μl of T7 RNA polymerase and 20 U of RNasin. Then, 2 μl SDS sampling buffer (125 mM Tris-HCl, pH 6.8, 120 mM 2-mercaptoethanol, 2 % SDS solution, 0.025 % bromophenol blue, 20 % glycerol) was added to 3 μl of the reaction solution, and the mixture was treated by heating at 95 °C for 3 minutes and subjected to SDS-polyacrylamide gel electrophoresis. By autoradiography, the molecular weight of the translated product was determined. As a result, the translation product, which had a molecular weight of 80 kDa almost similar to the molecular weight (80,618) deduced from the ORF, was formed.

(iv) Expression of GST fusion protein in *E. coli*

The translated region was amplified by PCR where pHP03494 was used as a template while a 26-mer sense primer (SEQ ID NO: 4) starting at a translation initiation codon and having an *Eco*RI recognition site added thereto and a 26-mer antisense primer (SEQ ID NO: 5) terminating at a termination codon having a *Sal*I recognition site added thereto were used respectively as primers. The PCR product was digested with restriction enzyme *Eco*RI and inserted into *Eco*RI site in vector pGEX-5X-1 (Pharmacia). After its nucleotide sequence was confirmed, the resulting plasmid was used for transforming *E. coli* BL21. The transformant was cultured at 37 °C for 5 hours in LB medium, and IPTG was added thereto at a final concentration of 0.4 mM, followed by culturing at 37 °C for 2.5 hours. The microorganism was separated by centrifugation and lysed in a lysing solution (50 mM Tris-HCl (pH 7.5), 1 mM EDTA-1 % Triton X-100, 0.2 % SDS, 0.2 mM PMSF), frozen once at -80 °C, thawed, and disrupted by sonication. After centrifugation at 1000 x g for 30

minutes, glutathione Sepharose 4B was added to the supernatant and incubated at 4 °C for 1 hour. After the beads were sufficiently washed, a fusion protein was eluted with an eluent (10 mM Tris-50 mM glutathione). As a result, a GST-HP03494 fusion protein having a molecular weight of about 110 kDa was obtained.

(v) Preparation of antibody

Domestic rabbits were immunized with the above fusion protein as the antigen to give antiserum. First, an antiserum fraction precipitating by 40 % saturation with ammonium sulfate was applied onto a GST affinity column to remove GST antibody. Then, the unadsorbed fraction was purified by a GST-HP03494-antigen column.

(vi) Western blotting

A lysate of human fibrosarcoma cell line HT-1080 was separated by SDS-PAGE, blotted onto a PVDF membrane, blocked for 1 hour at room temperature with 0.05 % Tween 20-PBS (TPBS) containing 5 % skim milk, and incubated with the antibody diluted 10,000-fold with TPBS. The sample was washed 3 times with TPBS and then incubated for 1 hour with horseradish peroxidase-labeled goat anti-rabbit IgG diluted 10,000-fold with TPBS. The sample was washed four times with TPBS and detected by luminescence with an ECL reagent (Amersham), to give a signal with a molecular weight of 80 kDa. This molecular weight agreed with the molecular weight of the *in vitro* translated protein product in the rabbit cell-free translation system.

(vii) Expression of GFP fusion protein

The translated region was amplified by PCR where pHP03494 was used as a template while a 26-mer sense primer (SEQ ID NO: 4) starting at a translation initiation codon having an *Eco*RI recognition site added thereto and a 26-mer antisense primer (SEQ ID NO: 5) terminating at a termination codon

having a *Sal*I recognition site added thereto were used respectively as primers. The PCR product was digested with restriction enzymes *Eco*RI and *Sal*I and inserted into *Eco*RI site in GFP fusion protein expression vector pEGFP-C2 (Clontech). After the nucleotide sequence was confirmed, HeLa cells were transfected by the lipofection method with the resulting plasmid pEGFP-C2-HP03494. Under a fluorescence microscope, the cells transfected with pEGFP-C2 showed fluorescence on the whole of the cells, whereas the cells transfected with pEGFP-C2-HP03494 showed fluorescence on their nuclei only. This result indicated that HP03494 is a protein present in nucleus.

(viii) Binding to a C-terminal domain (CTD) of RNA polymerase II

The translated region coding for WW domain was amplified by PCR where pHP03494 was used as a template while a 33-mer sense primer (SEQ ID NO: 6) starting at a translation initiation codon with a *Bam*HI recognition site added thereto and a 33-mer antisense primer (SEQ ID NO: 7) terminating at a termination codon with an *Eco*RI recognition site added thereto were used respectively as primers. The PCR product was digested with restriction enzymes *Bam*HI and *Eco*RI and then inserted into *Bam*HI-*Eco*RI sites in vector pGEX-5X-1 (Pharmacia). The resulting plasmid was subjected to expression in *E. coli* in the same manner as in (iv), to give a fusion protein GST-HP03494WW consisting of GST and HP03494 WW domain, and this fusion protein was separated by SDS-PAGE, then transferred onto a PVDF membrane, incubated with ^{32}P-labeled GST-CTD or ^{32}P-labeled GST-pCTD (GST-phosphorylated CTD) phosphorylated depending on a nuclear extract (Hirose, Y and Manley, J. L., Nature, 395, 93-96, 1998), and detected by the Far Western method (Kaelin, Jr. et al., Cell, 70, 351-364, 1992). It was revealed that the WW domain on HP03494 binds more strongly to phosphorylated CTD. This result suggested that the protein of this invention is involved in regulating transcription.

Industrial Applicability

This invention provides an isolated and purified human nuclear protein existing in human cell nucleus, a polynucleotide (human cDNA and genomic DNA fragment) encoding this protein, and an antibody against this nuclear protein. The protein and antibody of this invention are useful for diagnosis and therapy of morbid states such as cancers. By use of the present polynucleotide, the present protein can be expressed in a large amount. By screening a low-molecular compound binding to the present protein, a new type of pharmaceutical preparation such as antitumor agent can be searched for.